Age-neutrality of categorically and dimensionally measured DSM-5 Section II personality disorder symptoms

I. Debast, G. Rossi, S. P. J. van Alphen, E. Pauwels, L. Claes, E. Dierckx, H. Peuskens, E. Santens, & C. K. W. Schotte

Inge Debast and Gina Rossi

Department of Clinical and Lifespan Psychology, Vrije Universiteit Brussel, Belgium

S. P. J. (Bas) van Alphen

Department of Clinical and Lifespan Psychology, Vrije Universiteit Brussel (VUB), Belgium

Mondriaan Hospital, Heerlen-Maastricht, The Netherlands

Els Pauwels

Depertment of Clinical Psychology, Catholic University of Louvain

Alexian Brother's Hospital Tienen, Belgium

Laurence Claes

Department of Clinical Psychology, Catholic University of Louvain, Belgium

Eva Dierckx

Department of Clinical and Lifespan Psychology, Vrije Universiteit Brussel (VUB), Belgium

Alexian Brother's Hospital Tienen, Belgium

Hendrik Peuskens and Els Santens

Alexian Brother's Hospital Tienen, Belgium

Chris K. W. Schotte

Department Clinical and Lifespan Psychology, Vrije Universiteit Brussel (VUB), Belgium

Department Clinical Psychology, Universitair Ziekenhuis Brussel, Belgium

Abstract

Studies on the face validity of DSM-5 Section II categorical personality disorder (PD) symptoms indicate a bias against older adults. To extend these results, this paper explored whether categorically and dimensionally scored PD symptoms of DSM-5 Section II, as measured in the Assessment of DSM-IV Personality Disorders (ADP-IV; Schotte & de Doncker, 1994), corroborate potential age bias across younger (aged 18-34), middle aged (35-59 years) and older adults (aged 60-75). *Differential Item Functioning* (DIF) analyses, following a Classical Test Theory approach, showed that two of the 79 symptoms were measured differently across three age groups when categorically assessed, and four when dimensionally measured. Nevertheless, subsequent *Differential Test Functioning* (DTF) analyses supported a low aggregated impact of DIF on the dimensional scales, justifying mean-level comparisons across age groups. Generalizability of the results is discussed in light of methodological issues concerning the research of age-neutrality of PD symptoms, including the employed measurement instrument, PD symptom measurement approach, and sample and age range used to describe older adults.

Last decade, new insights in personality research with late-middle and older-age adults led to the conclusion that personality and personality disorders (PDs) are not as stable as they were assumed to be (see Debast et al.[2014] for a review of stability of personality traits and PDs in late-middle and old-age). However, to what extent actual change influences endorsing disorder related symptoms is less clear. The *heterotypical continuity* hypothesis argues that while the underlying latent characteristics remain the same, the disorder changes in its manifestation of related symptoms, instead of actual changes in the degree of PD severity (Mroczek, Hurt & Berman, 1999). Particularly, PDs can manifest differently in later life as a result of cognitive deterioration, somatic comorbidity, medication, and psychosocial challenges (van Alphen et al., 2012). This could mean that some items within PD instruments do not apply to older adults in the same way as they do to younger adults, as most of the items are built within the occupational, social, psychological or physiological context of younger adults. An obvious example is the Schizoid criterion "neither enjoys nor experiences sexual relations", which has little to do with Schizoid pathology in a later life context but rather refers to age-related physiological changes (van Alphen, Engelen, Kuin, & Derksen, 2006). Moreover, other criteria such as the Avoidant PD criterion, "avoids occupational activities" also refer explicitly to life domains of younger adults, as most of the older adults are retired (Tackett et al., 2009). Furthermore, according to a qualitative study by Dutch forensic psychiatrists and psychologists (Van Alphen, Nijhuis, & Oei, 2007), only three out of ten Antisocial PD features were deemed applicable to the elderly (65+): tends to justify behavior, shows dishonesty, and lacks remorse.

As the measured symptoms do not closely consider the presentation of personality in later life, the validity of the assessment of these disorders is possibly affected (Tackett, Balsis, Oltmanns, & Krueger, 2009). Oltmanns and Balsis (2011) consider the poor face validity of DSM PD symptoms the main cause of the limited understanding of the course of PDs in older adults. In addition, clinicians must be able to rely on available instruments, without having to adjust items to the context of their older patients (Zweig, 2008; Tackett et al., 2009). After all, the presence of PDs complicate treatment of clinical syndromes in all ages (van Alphen, Engelen, Kuin, Hoijtink, & Derksen, 2006; Zweig, 2008).

To make valid conclusions about the presence of PDs across different ages in the community, as well as in clinical samples, an age-neutral measurement instrument is clearly required (Balsis, Gleason, Woods, & Oltmanns, 2007; Tackett et al., 2009). Technically, an age-neutral instrument implies that items work the same for different age groups. One piece of evidence that contributes to age-neutrality of an assessment tool is invariance of item responses across age groups. Group invariance of item responses is accomplished when the item response, while having the groups matched on the latent or observed variable or scale, is the same and thus independent of group membership (Millsap & Everson, 1993). Matching the groups on the latent or observed scale score allows for the assessment of differential item scores by controlling for mean differences in the scale scores. So, if older and younger adults with the same PD level have different probabilities of endorsing particular items of that PD, *Differential Item Functioning* (DIF) occurs (Holland & Thayer, 1988). Violation of the invariance condition, in other words the presence of DIF, indicates that the item in question may measure not only construct-

relevant differences (i.e. PD-related) between two or more age groups but also other group characteristics related to age, which jeopardizes the age-neutrality and subsequently the validity of the age differences in PD scores (Balsis et al., 2007; Tackett et al., 2009).

Balsis and colleagues (2007) analyzed which DSM–IV PD criteria were hampered by DIF in a cross-sectional dataset of 18,565 community participants consisting of younger adults aged 18-34 years and older adults aged 65-98. Symptoms of seven DSM-IV PDs (Antisocial, Avoidant, Dependent, Histrionic, Obsessive–compulsive, Paranoid, and Schizoid) were assessed using the Alcohol Use Disorder and Associated Disabilities Interview Schedule DSM-IV version (AUDADIS-IV; Grant, Dawson & Hasin, 2001). This semi-structured interview measures the presence of pathological personality features on a categorical item scale. Consequently, they used the sum of PD symptoms endorsed to match the age groups on the PD level before responses on each item symptom were compared across age groups. The authors reported that 29% of the 79 symptoms were age-biased, influencing measurement across younger and older respondents to some degree for all investigated PDs.

However, the validity of symptom counts as an estimation of the PD level that is used as a matching or stratifying variable in PD has been criticized recently. In contrast to a categorical approach of symptom assessment only yielding a coarse estimation of each PD level, dimensionally measured symptoms allow a more fine-grained measurement of PD levels (Cooper, Balsis, & Zimmerman, 2010) and follow the new assessment approach of PD traits implemented in Section III of the DSM-5. Van den Broeck, Bastiaansen, Rossi, Dierckx and De Clercq (2013) investigated potential age bias of the

Section III trait dimensions measured by the Personality Inventory for DSM-5 (PID-5; Krueger, Derringer, Markon, Watson, & Skodol, 2012). In this trait model, 25 primary pathological trait facets are organized by five higher-order domains (Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism) to conceptualize six types of PDs (Schizotypal, Antisocial, Borderline, Narcissistic, Avoidant, an Obsessive-compulsive) (APA, 2013). The authors found by means of DIF analyses that 15% of the 220 items may contain age-related DIF in a community sample. In order to be able to compare DIF in dimensional versus categorical measurement of PD symptoms, we will apply the Assessment of Personality Disorders of the DSM-IV (ADP-IV; Schotte & de Doncker, 1994) in the current study. The ADP-IV is a self-report instrument that allows for both a dimensional and categorical assessment of the DSM-5 Section II PD symptoms. Moreover, we also address other shortcomings of previous research. Firstly, as the Borderline, Narcissistic and Schizotypal PDs were not measured in the age neutrality study of section II PDs in Balsis et al. (2007), all Section-II PD symptoms will be measured using the ADP-IV. Secondly, we will examine age-neutrality of PDs in a clinical sample, since former studies were limited to community samples. Thirdly, the inclusion of a large age range in the older adult group could have influenced the results in Balsis et al. (2007) and Van den Broeck et al. (2013), where older adults ranged from 65 to 98 and from 61 to 99 respectively. Life experiences, physical challenges, psychological experiences, and social opportunities can differ between the age from 65-75 (referred to as the "young-old"), 75-85 ("old-old") and 85 through older (the "oldest-old") (Segal, Coolidge, & Rosowsky, 2006), resulting in a heterogeneous population. Research on age-neutrality of PD symptoms needs to be fine-tuned by focusing on

specific age categories when exploring the patterns of DIF in older adults. Furthermore, examining more than two successive age groups will broaden insight in patterns of DIF. Besides item level analyses, it is finally critical to test for the aggregated effect of present DIF at the scale level to determine for which scales the mean PD scores can be justifiably compared among age groups, and thus can be validly measured across age groups by clinicians in practice.

Method

Sample

Data were collected in a residential unit for alcohol and drug treatment, located in a psychiatric hospital in Flanders. At admission, all patients provided a written informed consent to use their clinical assessment records for scientific purposes. Questionnaires were completed during the first two weeks after admission. Patients in an acute state of mental crisis, and patients who could not complete the questionnaire due to cognitive limitations were not included in this study. Data collection was established using partly paper and pencil measures and partly computerized administrations. The study was approved by the ethical committee of the hospital.

Consecutive patients in the clinical records between 2008-2013 with a minimum age of 60 who completed ADP-IV questionnaires were included in the study. The comparison groups of younger and middle aged adults were matched with records from the same database to the group of older adults based on sample size and gender rates to obtain homogenous groups. In total, a sample of 321 adults was analyzed from a database

of approximately 1500 patients, comprising three age groups of each 107 participants[1]. The 107 "older adults" ranged from 60 to 75 years (*M* = 63, *SD* = 3.30). We evaluated 18-34 year olds (*M* = 28.68, *SD* = 4.15) to whom we refer as "younger adults", which is also the applied age range in Balsis et al. (2007). We also evaluated a "middle aged" group comprised of individuals between ages 35-59 (*M* = 47.26, *SD* = 6.93). All age groups consisted of 62% males and 38% females.

Supplemental information about the clinical characteristics of the age groups, such as the nature of substance abuse disorder and other clinical symptoms, was obtained using the Drug Use Screening Inventory – Revised (DUSI-R; Tarter & Kirisci, 2001) and the Symptom Checklist-90 (SCL-90; Arindell & Ettema, 1986). The participants were categorized in three groups according to the primary substance of use, namely alcohol, illicit drug, and dual-usage (i.e. co-occurrence of alcohol and illicit drug use).

Chi-square- and *F*-statistics were used to evaluate the association between age groups and the clinical characteristics. Post-hoc tests examined the significance of differences in the nature of substance abuse and Symptom Checklist -90 (SCL-90; Arrindell & Ettema, 1968) subscales with pairwise comparisons and Bonferroni adjusted *p*-values adjusted for the number of comparisons made in each analysis ($p<.016$; .05 divided by 9 comparisons for 3 categories of substance abuse x 3 age groups, and $p<.006$; .05 divided by 3 comparisons for 3 age groups in each F-test of the SCL-90 subscales). The older adults endorsed less dual usage than the younger adults with a small

[1] These participants were also enlisted by Pauwels et al. (2014) for the examination of DIF in the Dutch version of the Young Schema Questionnaire long form (YSQL2; Dutch version: Young & Pijnaker 1999).

effect according to Cohen's *d* effect sizes (Cohen, 1988). On the other hand, the younger adults used less alcohol than the older adults, as evidenced by a medium effect (see Table 1). As for clinical psychopathology, no significant differences were found, except for a small effect in the SCL-90 subscale Hostility, with on average higher scores for the younger adults in comparison to the middle aged and older adults (see Table 2).

Instruments

The *Assessment of DSM-IV Personality Disorders* (ADP-IV; Schotte & de Doncker, 1994) is a self-report instrument that consists of 94 items, representing the symptoms of the 10 Section II DSM-IV PDs criteria and the research criteria of the Depressive and Passive-Aggressive PD. For each PD symptom, the ADP-IV assesses the self-judged typicality by means of a seven-point 'Trait scale', ranging from 1 (*Totally disagree*) to 7 (*Totally agree*). When judged as typical (score 5, 6 or 7), the degree of dysfunction the symptom has caused for the person or others is assessed by means of a three-point 'Distress-scale', ranging from 1 (*Totally not*) to 3 (*Most certainly*). The ADP-IV allows for two diagnostic assessment formats. First, dimensional PD scores result from summing the Trait-scores for each PD scale. Norms for men and women were developed in a stratified sample of the Flemish general population (n = 659) and serve as a guide for the interpretation of the dimensional scores (Schotte, 2000). Second, a categorical diagnosis requires a Trait score > 4 and a Distress score >1 for the number of symptoms that is needed to meet a disorder. Previous research with the ADP-IV (Schotte, de Doncker, Vertommen, Vankerckhoven, & Cosyns, 1998) revealed that the dimensional Trait-scales are internally consistent (median Cronbach's alpha: .76; range: .60-.84) in a stratified sample of the Flemish population. The ADP-IV's short-term

and long-term reliability have been documented as well (Schotte, 2000). Evidence for the construct validity of the categorical diagnoses and dimensional scales has been reported, and the ADP-IV has been translated into several languages, including English (Schotte et al., 2004).

Analyses

The potential age bias of the items across age groups was examined using Differential Item Functioning (DIF) analyses. Technically, an item is said to exhibit DIF if two groups with a similar position on a particular scale (e.g., Avoidant PD) do not have the same probability of endorsing an item. Current analyses were conducted using an odds ratio approach from Classical Test Theory (CTT). In contrast to Item Response Theory, this non-parametric method has the advantage that it can be used with smaller sample sizes and does not require model fit (Zumbo, 1999). DIF analyses were conducted using the freely available DIFAS 4.0 program (Penfield, 2007). The odds ratio or contingency approach of DIF analysis in this program tests whether the item score is higher (or rather, the odds for a high item score is greater) in one group compared to another group with the same PD score, for all levels of the scale score, also called *uniform DIF*. In other words, it measures the overall effect of item-level invariance across the PD scale (Penfield, 2010).

To examine DIF across the three age groups, three pairwise DIF analyses were performed for each PD at the categorical and dimensional PD symptom measurement level of the ADP-IV. Only the typicality of each symptom as measured by the Trait-aspect is considered in these analyses, therefore leaving out the Distress-aspect in the measurement of the symptoms. In the DIF analyses for categorical or 'dichotomous'

items, the amount of endorsed symptoms in each PD based on T>4 algorithm in the ADP-IV was used as a conditional *(stratifying)* variable, representing the observed PD level on which the age groups are matched. The stratum size, or the unit of the stratifying variables, was set to one for both the categorical and dimensional DIF-analyses, which is the default setting in DIFAS 4.0.

The Mantel-Haenszel (MH) statistic (Holland & Thayer, 1988), a chi-square distribution with one degree of freedom (Penfield, 2007), is used to detect statistically significant DIF. Furthermore, two effect size indicators were examined (cfr. Guilera, Gomez-Benito, Hidalgo, & Sanchez-Meca, 2013). The Mantel-Haenszel common log-odds ratio (MH LOR; Camilli & Shepard, 1994) estimates the population ratio of the odds of item endorsement of the *reference group* (older adults) over that of the *focal group* (either the younger or middle aged adults) conditional on ability (Penfield & Algina, 2003). Positive values correspond to DIF "favoring" the reference group, which means that the reference group is more likely to endorse a higher item score than the focal group despite the same level of PD symptoms, and negative values indicate DIF "in favor of" the focal group (Penfield, 2007). The Educational Testing Service (ETS; Zieky, 1993) Scheme, based on a transformation of the Mantel Haenszel common log odds ratio, categorizes dichotomous items as having negligible (A), moderate (B), and high (C) levels of DIF.

In the DIF analyses for dimensionally measured or *polytomous* items, the ADP-IV summated Trait scores in each PD scale were used as stratifying variables, representing the degree to which the PD existed in each age group. The Mantel Chi-square (Mantel) statistic (Zwick, Thayer & Mazzeo, 1997) was used as a first step in the process of

detecting DIF. In addition, two alternative estimates of polytomous DIF were consulted to estimate the direction and effect size of the DIF: the Liu-Agresti Log Odds Ratio (L-A LOR; Liu & Agresti, 1996) and Cox's Noncentrality Parameter Estimator (Cox's B; Camilli & Congdon, 1999). The L-A LOR statistic delivers an estimate of the ratio of the reference group odds to obtain an item score as opposed to the focal group. The L-A LOR statistic as an estimator of an odds ratio provides a measure of effect size that is relatively independent of the number of response categories in an item (Penfield & Algina, 2003). The Cox's B is the sum of the odds ratio of all response options together (see Camilli and Congdon (1999). Similar to the MH LOR for dichotomous items, positive values in the L-A LOR and Cox's B indicate DIF "in favor of" the reference group (older adults), whereas negative values point to DIF "against" the reference group.

To examine the extent in which the measurement of PD dimensions across age groups is influenced by the amount of DIF variance in the items, the impact of DIF variance at the PD scale levels is estimated using Differential Test Functioning analyses (DTF) with the corresponding ADP-IV dimensional scales as stratifying or conditioning variables. Effect sizes are interpreted as small for $v^2>0.07$, medium for $0.07<v^2<0.14$ and large for $v^2>0.14$ (Penfield & Algina, 2006).

In order to reduce possible Type I inflation of incorrectly identifying items as exhibiting DIF, we adjusted the critical values of the DIF statistics across all items in each PD scale and the amount of test statistics. We applied a stringent Bonferroni corrected critical Mantel and MH value, ranging from 9.23 to 9.89, depending on the number of items in each scale. For example, the Bonferroni correction in the Paranoid scale across seven items and three test statistics was calculated as follows: 0.05/(7*3).

Similar stringent cut-off criteria were derived for the L-A LOR and MH LOR (ranging from 0.90 to 0.94) and Cox's B (ranging from 0.57 to 0.59) statistics to flag items with large DIF. As a result, the current statistical approach aims at identifying items showing large DIF. Only one of the two effect sizes parameters has to indicate a large DIF effect because of the stringent DIF detection procedure. A Bonferroni correction was applied for the DTF statistic as well, by dividing the cut-off score of 0.14 for large DIF variance through the critical value in LA LOR of 0.43 as proposed by Penfield and Algina (2006), multiplied by the adjusted LA-LOR (also see Van Den Broeck et al., 2013). In the current study, DTF critical values ranged from 0.35 to 0.36.

Regarding descriptive statistics, the Cronbach's alpha coefficients for each PD and age group were consulted to ensure reliable DIF results. A poor internal consistency in older adults compared to younger adults may indicate that some of the items measure a different construct in that age group (Balsis, Segal & Donahue, 2009). Finally, in order to explore possible differences in PDs between the age groups and to grasp for possible change along the life course, mean PD scales for which no large DTF was found, were compared across the age groups, by means of Analyses of Variance (ANOVA) in SPSS© version 22, along with effect sizes. Ten ANOVA's were conducted with each PD scale as dependent variable and age category as independent variable. Post-hoc tests were conducted to find out which age groups differ. A Bonferroni correction was applied to control the type I error rate (p<.016).

Results

Missing values

In DIFAS 4.0, cases for which the stratifying variable is missing are omitted from the entire analysis by means of listwise deletion. The proportion of missing values ranged from 0.6% for the Antisocial PD to 3.1% for the Paranoid PD. The probability of observation of missing data was associated with the age category. More paper and pencil administrations were apparent in the older adult group then the younger age groups, allowing items to be left open in contrast to the computerized administrations. For the effectiveness of the contingency method for measuring DIF, one should attempt to ensure that the sample sizes of groups being compared are not too disproportionate, in particular in small sample sizes (Guilera et al., 2013). Therefore, a regression-based imputation approach was performed in SPSS to control for the impact of missing data of the DIF-analyses. In this missing data treatment, values were predicted based on existing relations between all present variables. Some constraints were set on the imputed values. The minimum and maximum values were set to one and seven respectively, corresponding to the response options in the ADP-IV. Accordingly, because DIFAS only stratifies on integer values, integer values were requested via the rounding option.

Differential item and test functioning

To examine the possible presence of DIF at the categorical measurement level of the ADP-IV items across three age groups, a total of 30 separate analyses were conducted making pairwise comparisons between the three successive age groups for 10 PDs. Only the symptoms of the former 10 PDs that are included in the DSM-5 Section II were used for this study, leaving the Passive-Aggressive and Depressive PDs out of consideration. According to the stringent critical value of the MH statistic, two of the 79 items displayed significant DIF. More specifically, one Schizoid and one Dependent criterion showed

meaningful DIF in favor of the older adults compared to the younger adults, indicating they were more readily endorsed by the older adults than the younger adults despite equal amount of endorsed Schizoid and Dependent PD symptoms. Table 3 presents the paraphrased content and statistical values of these PD symptoms with large DIF concerning categorical measurement.

To investigate the possible presence of DIF at dimensional item measurement level across three age groups, again 30 separate analyses were performed. Table 4 illustrates the paraphrased content and statistical values of items with large DIF, along with the PD scale they belong to. The Bonferroni adjusted LA-LOR confirmed large DIF for four items, belonging to Schizoid, Antisocial, and Dependent PDs. One of these items showed DIF against the oldest sample, as opposed to the middle-aged adults, indicating this Antisocial item was more readily endorsed by middle-aged adults compared to older adults, after controlling for levels of the underlying level of Antisocial PD symptoms. In contrast, one item showed DIF in favor of the older adults when compared with the middle-aged adults and younger adults, indicating this item was more readily endorsed by the older adults despite equal levels of Dependent PD symptoms. The remaining DIF item, belonging to the Schizoid PD, was endorsed to a higher degree by the older adults in comparison with the younger adults while matched on their Schizoid PD symptom level.

To evaluate the impact of these items at scale level, the presence of DTF was analyzed. According to the scale-specific Bonferroni cut-offs, DTF was absent in every PD.

Descriptive statistics

In the current sample, Cronbach's alpha coefficients of the ADP-IV dimensional PD scales varied from .68 (Schizoid PD scale) and .83 (Avoidant PD scale), indicating good internal consistency of the dimensional Trait-scales. These alpha reliabilities are in line with findings of Schotte et al. (1998). For the three age groups separately, Cronbach's alphas ranged from .67 to .84 for the younger adults, from .58 to .86 for the middle aged adults and from .50 to .78 in the older age sample. Table 5 presents the Cronbach alpha coefficients, mean values, and standard deviations for each ADP-IV PD scale within the age groups, together with ANOVA's and associated effect sizes shown for the pairwise comparisons between the three age groups. Cross-sectional trends of mean values were investigated for PD scales in which there was no significant DTF found. As a result, all scales were analyzed regarding to their age-related mean differences. Analysis of each dependent variable, using a Bonferroni adjusted alpha level of .016 showed there was a significant effect for the Paranoid, Schizoid, Schizotypal, Antisocial, Borderline, Histrionic, Narcissistic and Avoidant PD scales. Post-hoc comparisons showed that, compared to older adults, the younger adults scored significantly higher on all these PD scales except the Schizoid PD scale, where the opposite occurred. In addition, the younger adults scored significantly higher than the middle aged group on the Paranoid, Antisocial and Borderline PD scales.

Discussion

This study examined the potential age bias of categorical and dimensional measurement of DSM-5 Section II PD symptoms, by means of differential item

functioning tests within a CTT framework and the same instrument, the ADP-IV. DIF occurred when the younger and the older adults matched with respect to their degree of PD pathology had a differential probability of endorsing any given symptom criterion. Categorically measured PD symptoms were found to show DIF in only two of the 79 symptoms, compared to four symptoms when measured dimensionally. More specifically, one Schizoid and one Dependent item were more readily endorsed by the older adults than the younger adults, despite a similar degree of corresponding PD symptoms. The Schizoid criterion "Having no interest in sexual encounters with someone else is characteristic of me" was endorsed at a higher degree by the older adults in comparison to both younger age groups with the same level of Schizoid PD, in line with earlier intuitive and empirical findings. Likewise, the Dependent criterion "I find it very difficult to disagree in public; even if I totally disagree with someone I don't dare to voice my opinion and therefore will agree" exhibited DIF in favor of the older adults. These symptoms also displayed DIF when measured dimensionally, in the same way as in the categorical scores. At the same time, a dimensional DIF effect of the Dependent symptom in favor of the older adults also was found in relation to the middle aged adults. This trend suggests that "dependent" older adults have more difficulty disagreeing with others then their younger counterparts. This supports DIF in this criterion as found by Balsis et al. (2007), but adds to the former evidence that this pattern also exists when relating to the middle-aged adults. Beside this PD symptom, an additional DIF item appeared when measured dimensionally. The Antisocial PD criterion "Not meeting my responsibilities and obligations (whether they are financial, professional, or in caring for my family …) is typical of me" was expressed in a lower degree by the older adults as

opposed to the younger adults with the same Antisocial PD score. The Antisocial PD scale is also marked by a poor internal consistency reliability in the older adult group as indicated by a Cronbach's alpha coefficient of .50. Further examination of the low alpha value revealed that the majority of the Antisocial PD items in the ADP-IV may be more differentially related to the Antisocial PD construct in the older as opposed to younger adults. Therefore, the interpretation of this DIF needs caution.

In short, the current results support the hypothesis that dimensional PD items provided a more nuanced analysis of different PD symptom presentation across age groups. As these dimensional items were not evidenced as large DIF by the Cox's B statistic as by the L-A LOR statistic, the items represent a DIF-effect that is cumulative across all response categories in that item but not tied to a particular item score. Nevertheless, the results indicate that, except for four symptoms, the vast majority of the dimensionally measured items were similarly endorsed by the younger, middle aged and older adults with the same score on the PD scale. Above all, the symptoms were always found to function equally between the two youngest age categories (18-34 versus 35-59).

Altogether, there was a relatively small amount of DIF in comparison with previous research. Although it was expected that the IRT method as conducted by Balsis and colleagues (2007) is more sensitive than the CTT method by extending the analyses to non-uniform DIF detection, our oldest age group with age ranging from 60 to 75 did not include the so called "old-old" (aged 75-85) and "oldest old" (aged 85 through older) (c.f., Segal et al., 2006). Consequently, the results do not permit generalization to the heterogeneous elderly population as a whole as much as in the previous studies. Moreover, the demarcation age of 65 for older age is used by Balsis and colleagues

(2007). Life experiences, physical challenges, and psychological experiences could have influenced our results, as many older adults in this study did not reach the age of retirement at that point. Apart from the sample characteristics, the ADP-IV has an age-neutral measurement intention by referring in its objective to characteristics that are present from early adulthood and show up in a large number of personal and social situations. In the instructions, one is asked to judge each statement based on one's own life's experience. Hence, problems with face validity seem to be compromised by a generalization and evaluation of personality characteristics over the lifespan.

By consequence, the aggregated DIF variance at the scale level of the PDs, measured through Differential Test Functioning (DTF) analyses, was minimal for all PD scales. In other words, the amount of DIF in the items did not result in whole PD scale scores being biased. This implicates that, in spite of some items containing DIF, comparisons of mean PD scores between the younger, middle-aged, and older adults are justified for all ten PD scales. Considering the mean differences across the three age groups, the Paranoid, Schizotypal, Antisocial, Borderline, Histrionic, Narcissistic, and Avoidant PD scales diminish significantly from young adulthood to the age of 75, with significant decline in the Paranoid, Antisocial and Borderline PD scales already apparent in middle age. An exception is found in the Schizoid PD scale, where the older adults score significantly higher as opposed to younger adults. The mean differences in the Schizoid and Cluster B disorders are in line with other cross-sectional and longitudinal findings (Debast et al., 2014). For example, to understand the remission in externalizing personality traits characterizing the Cluster B PDs, Roberts, Walton and Viechtbauer (2006) explained that older adults show less impulsivity and risky, irresponsible behavior

because of changes in physical condition accompanying aging and subsequent limited mobility. The decline in the Paranoid, Schizotypal and Avoidant PD scales in older adults, in comparison to younger adults, was possibly also influenced by the differences in the primary substance of abuse: The younger adults endorsed more dual dependency, and less dependency only on alcohol than older adults. Colpaert, Vanderplasschen, De Maeyer, Broekaert and De Fruyt (2012) found similar differences in ADP-IV dimensional PD scales between patients dependent on alcohol only, drugs only, and alcohol and drugs. It is important to note that, despite mean differences in the ADP-IV PD scales, there is low overall DIF variance throughout the items. Therefore, these differences across age groups in the substance use setting can be considered reliable.

Limitations

Although the present study has managed to explore age-related DIF in three small clinical age groups with the contingency method by using equal sample sizes, some methodological limitations should be discussed when interpreting the results. The uniform and net DIF approach that is characteristic for the Classical Test Theory approach existed in detecting consistent DIF across all scale and item levels. Due to mean differences in the PD dimensional scales across age groups, some values in the conditioning variable had frequencies of zero for one of the age groups. As a result, these values were excluded from the DIF-analyses. Concerning the dichotomous DIF-analyses, the highest levels of symptoms throughout all ten PDs were never endorsed by the middle-aged or the older adults. Nevertheless, all five norm ranges of the stratified norm group of the Flemish general population (very low, low, average, high, very high) for the ADP-IV dimensional scales were represented, allowing the dimensional DIF-findings to

generalize across very low to very high PD pathology scores. In addition, the cross-sectional design makes it difficult to distinguish between age or cohort effects in the interpretation of the results. Beyond the influence of age and cohort, other demographic differences also cannot be ruled out. Similarly, all our participants were hospitalized for substance dependency, yet they differed in the primary substance of use. Since we could not match the groups on this variable, confounds may exist between age group and type of substance use. Although the use of comparable groups is difficult to accomplish, efforts to obtain groups as homogenous as possible is warranted to detect age bias exclusively. However, it was not our intention to examine the cause of the current DIF results, but simply to detect DIF as a function of age in a substance use setting. On the whole, the ADP-IV functioned equally in the substance disorder setting across age groups, despite the somewhat different clinical and perhaps other characteristics. The dimensional scoring format of the ADP-IV seems to be promising for assessing the DSM-5 Section II PDs in older adults aged 60 to 75 in a substance use population.

In order to determine if the DIF presents item bias and what causes it, one would need to apply more sophisticated IRT-based item-bias analyses, as Balsis and colleagues (2007) already have mentioned. As research regarding age-neutrality is limited to cross-sectional findings so far, a longitudinal investigation is recommended. Also, inclusion of more various and successive older age groups, with larger marginal sample sizes, are needed to refine the current findings. Moreover, it would be useful to study the age-neutrality of the ADP-IV in patients with different kinds of psychopathology, such as in- and outpatients with depression or anxiety disorders.

In addition, the current and previous research relied exclusively on self-report instruments. In geriatric psychiatry, informant reports are often indispensable due to cognitive problems and related impaired insight and judgment (Barendse, Thissen, Rossi, Oei & van Alphen, 2013). Therefore, it would be interesting to investigate the age-neutrality of PD symptoms through informant report. Finally, as dimensional measurements also are more capable of predicting personality dysfunction (Skodol et al., 2011), it is of interest to explore the relations between dimensional pathological personality traits and experienced personality dysfunction across the lifespan, the latter aspect being an important condition to determine a PD diagnosis. Different weights could be determined for the symptoms to demarcate their presence depending on the age group to which the respondent belongs in order to avoid under- or overdiagnosis, as demonstrated by Cooper et al. (2010), rather than weighting each criterion equally for younger and older adults as is being done up to the current DSM.

Because the reported findings are exploratory and still suggestive, more expansive research on the possibilities of dimensional constructs of PDs is recommended. Nevertheless, this study is an important step in stimulating research on the age-neutrality of DSM-5 PDs.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013*). Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Arrindell, W.A., & Ettema, J.H.M. (1986). *SCL-90. Een multidimensionele psychopathologie indicator*. Lisse: Swets & Zeitlinger.

Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging, 22,* 171–185.

Balsis, S., Segal, D. L., & Donahue, C. (2009). Revising the Personality Disorder Diagnostic Criteria for the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM- V): Consider the Later Life Context. *American Journal of Orthopsychiatry*, 79, 452-460.

Barendse, H.P.J., Thissen, A.J.C, Rossi, G., Oei, T.I., & Alphen, S.P.J. van. (2013). Psychometric properties of an informant personality questionnaire (the HAP) in a sample of older adults in the Netherlands and Belgium. *Aging and Mental Health, 17,* 623-629.

Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics, 24,* 323–341.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.).New Jersey: Lawrence Erlbaum Associates.

Colpaert, K., Vanderplasschen, W., De Maeyer, J., Broekaert, E., & De Fruyt, F. (2012). Prevalence and determinants of personality disorders in a clinical sample of alcohol-, drug-, and dual-dependent patients. *Substance Use & Misuse, 47,* 649-661.

Cooper, L.D., Balsis, S. & Zimmerman, M. (2010). Challenges Associated With a Polythetic Diagnostic System: Criteria Combinations in the Personality Disorders. *Journal of Abnormal Psychology, 119,* 886–895.

Debast, I., van Alphen, S.P.J., Tummers, J.H.A., Rossi, G., Bolwerk, N., Derksen, J.L.L., & Rosowsky, E. (2014). Personality traits and personality disorders in late middle and old age: do they remain stable? A literature review. *Clinical Gerontologist, 37,* 253-271.

Grant, B.F., Dawson, D.A., & Hasin, D.S. (2001). *The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version*. National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.

Guilera G., Gómez-Benito J., Hidalgo M.D., & Sánchez-Meca J. (2013). Type I error and statistical power of the Mantel-Haenszel procedure for detecting DIF: a meta analysis. *Psychological Methods, 18,* 553-571.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial

construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42,* 1879–1890.

Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratio with a stratified ordinal response. *Biometrics, 52,* 1223–1234.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17,* 297–334.

Mroczek, D. K., Hurt, S. W., & Berman, W. H. (1999). Conceptual and methodological issues in the assessment of personality disorders in older adults. In E. Rosowsky, R. C. Abrams, & R. A. Zweig (Eds.). *Personality Disorders in Older Adults* (1st ed., pp. 135–150). Mahwah: Lawrence Erlbaum Associates.

Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 7,* 321–349.

Pauwels, E., Claes, L., Dierckx, E., Debast, I., van Alphen, S.P.J., Rossi, G., Schotte, C., Santens, E., & Peuskens, H. (2014). Age neutrality of the Young Schema Questionnaire in patients with a substance use disorder. *International Psychogeriatrics,* doi 10.1017/S1041610214000519.

Penfield, R. D. (2007). *DIFAS 4.0. Differential item functioning analysis system: user's manual.* Unpublished manuscript.

Penfield, R. D. (2010). Distinguishing between Net and Global DIF in Polytomous Items. *Journal of Educational Measurement, 47,* 129-149.

Penfield, R. D., & Algina, J. (2003). Applying the Liu-Agresti estimator of the

cumulative common odds ratio to DIF detection in polytomous items. *Journal of Educational Measurement, 40,* 353-370.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43,* 295–312.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change In personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 3–27.

Schotte, C. K. W., & de Doncker, D. (1994). *ADP-IV Vragenlijst.* Antwerpen: Universitair Ziekenhuis Antwerpen.

Schotte, C. K. W., & de Doncker, D. (2000). De ADP-IV: Een vragenlijst voor een therapeutisch georiënteerde diagnostiek van persoonlijkheidsstoornissen [The ADP-IV: A questionnaire for therapeutic oriented diagnostics of personality disorders]. *Psychopraxis, 2,* 267–273.

Schotte, C. K. W., de Doncker, D.A.M., Dmitruk, D., Van Mulders, I., D'Haenen, H., & Cosyns, P. (2004). The ADP-IV questionnaire: differential validity and concordance with the semi-structured interview. *Journal of Personality Disorders, 18,* 405-419.

Schotte, C. K. W., de Doncker, D., Vankerckhoven, C., Vertommen, H., & Cosyns, P. (1998). Self-report assessment of the DSM-IV personality disorders: Measurement of trait and distress characteristics – The ADP-IV. *Psychological Medicine, 28,* 1179–1188.

Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2006). *Personality Disorders and Older Adults: Diagnosis, Assessment, and Treatment.* Hoboken, NJ: Wiley.

Skodol, A.E.,.Clark, L. A., Bender, D.S., Krueger, R.F., Morey, L.C., Verheul, R., Alarcon, R.D., Bell, C.C., Siever, L.Y., & Oldham, J.M. (2011). Proposed changes in personality and personality disorder assessment and diagnosis for *DSM-5* Part I: Description and rationale. *Personality Disorders-Theory Research and Treatment, 2,*4-22.

Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span: developmental considerations for the fifth edition of the diagnostic and statistical manual of mental disorders. *Development and Psychopathology, 21,* 687–713.

Tarter, R. & Kirisci, L. (2001). Validity of the Drug Use Screening Inventory for predicting DSM-III-R substance use disorder. *Journal of Child & Adolescence Substance Abuse. 10,* 45–53.

van Alphen, S.P.J. Bolwerk, N., Videler, A.C., Tummers, J.H.A., Van Royen, R.J.J., Barendse, H.P.J., Verheul, R., & Rosowsky, E. (2012). Age related aspects and clinical implementations of diagnosis and treatment of personality disorders in older adults. *Clinical Gerontologist, 35,* 27–41.

van Alphen, S.P.J., Engelen, G.J.J. A., Kuin, Y., Hoijtink H.J.A., & Derksen, J.J.L.(2006). A preliminary study of the diagnostic accuracy of the Gerontological Personality Disorders Scale (GPS). *International Journal of Geriatric Psychiatry*, *21*, 862–868.

van Alphen, S. P. J., Nijhuis, P. E. P., & Oei, T. I. (2007). Antisocial personality disorder

In older adults. A qualitative study of Dutch forensic psychiatrists and forensic psychologists. *International Journal of Geriatric Psychiatry, 22,* 813-815.

van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., & Derksen, J. J. L. (2006). The relevance of a geriatric sub-classification of personality disorders in the DSM-V. *International Journal of Geriatric Psychiatry, 21, 2*05–209.

Van den Broeck, J., Bastiaansen, L., Rossi, G., Dierckx, E., & De Clercq, B. (2013). Age neutrality of the trait facets proposed for personality disorders in DSM-5: A DIFAS analysis of the PID-5. *Journal of Psychopathology and Behavioral Assessment, 35,* 487-494.

Young, J. E., & Pijnaker, H. (1999). *Cognitieve therapie voor persoonlijkheidsstoornissen: Een schemagerichte benadering [Cognitive therapy for personality disorders: A schema oriented approach]*. Houten, The Netherlands: Bohn Stafleu Van Loghum.

Zieky, M. (1993). Practical questions in the use of DIF statistics in item development. In P. W.Holland & H. Wainer (Eds.). *Differential Item Functioning* (pp. 337–347). Hillsdale, NJ: Lawrence Erlbaum.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zweig, R. A. (2008). Personality disorder in older adults: assessment, challenges and strategies. Professional Psychology: *Research and Practice, 39,* 298–305.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures

for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*, 321-334.

Table 1

Descriptive statistics, Chi-square and effect sizes for prevalence of alcohol and substance use for the younger (n=99) middle aged (n=104) and older (n=67) adults.

	Valid percentages			Chi-square test		Effect size Cohen's *d*		
Substance use	Young Adults	Middle Aged Adults	Older Adults	$X^2(2,270)$	*p*-value	*1 vs. 2*	*2 vs. 3*	*1 vs. 3*
Alcohol use	12%	21%	39%	16.574	.000*	.24	.39	-.66**
Illicit drug use	8%	3%	3%	3.639	.162	.23	.01	.21
Dual usage	80%	76%	58%	10.212	.006*	.09	.38	.48**

Note. Effect sizes were based on comparisons of the younger versus middle aged group 1 vs. 2, middle aged versus older adults 2 vs. 3, and younger versus older adults 1 vs. 3; *significant X^2 at $p < .05$; **pairwise comparisons at $p < .006$.

.

Table 2. *Descriptive Statistics, Anova and Effect Sizes for the SCL-90 Scales for the Younger (n=99), Middle-Aged (n=104), and Older Adults (n=67)*

	Raw Means (SD)			Analysis of Variance		Effect size Cohen's *d*		
SCL-90 Subscale	Younger	Middle aged	Older	*F* (df=2,270)	*p*-value	1 vs. 2	2 vs. 3	1 vs. 3
Anxiety	20.26 (8.15)	19.95 (7.91)	18.21 (8.29)	1.981	.140	.04	.22	.25
Agoraphobia	10.06 (4.28)	10.06 (4.25)	9.97 (4.925)	.012	.988	.00	.02	.02
Depression	36.96 (13.58)	36.96 (1.28)	37.06 (14.92)	.001	.999	-.00	-.01	-.01
Somatising problems	20.70 (8.25)	21.08 (7.32)	20.95 (7.77)	.067	.935	-.05	.02	-.03
Insufficiency of thought and behavior	18.96 (7.22)	19.40 (7.27)	18.08 (6.53)	.972	.380	-.06	.19	.13
Distrust and interpersonal sensitivity	32.65 (11.38)	33.72 (10.83)	32.31 (12.17)	.437	.646	-.10	.12	.03
Hostility	9.52 (3.79)	8.41 (3.09)	8.33 (2.79)	4.492	.012*	.32**	.03	.36**
Sleep problems	7.43 (3.52)	7.05 (2.97)	7.42 (3.57)	.455	.635	.12	-.11	.00

Note. Effect sizes were based on comparisons of the younger versus middle aged group 1 vs. 2, middle aged versus older adults 2 vs. 3, and younger versus older adults 1 vs. 3; *significant F at $p < .05$; **significant pairwise comparisons at $p < .016$.

Table 3

Categorically measured symptoms meeting Bonferroni adjusted cut-offs for large DIF

PD scale	Item	Paraphrased item content	MH	MH LOR	ETS
Items displaying DIF in favor of older adults compared to younger adults (older > younger)					
Schizoid	26	Having no interest in sexual encounters with someone else is characteristic of me	10.69	1.44	C
Dependent	33	I find it very difficult to disagree in public; even if I totally disagree with someone I don't dare to voice my opinion and therefore will agree	17.34	2.00	C

Note. MH=Mantel-Haenszel Chi-square; MH LOR= Mantel-Haenszel Log Odds Ratio; ETS= Educational Testing Service scheme; C=Large effect size.

Table 4

Dimensionally measured items meeting Bonferroni adjusted cut-offs for large DIF

PD scale	Item	Paraphrased item content	Mantel	L-A-LOR	Cox's B
Items displaying DIF against older adults compared to middle aged adults (older < middle aged)					
Antisocial	64	Not meeting my responsibilities and obligations (whether they are financial, professional, or in caring for my family …) is typical of me.	11.25	-1.20	-0.37
Items displaying DIF in favor of older adults compared to middle aged adults (older > middle aged)					
Dependent	33	I find it very difficult to disagree in public; even if I totally disagree with someone I don't dare to voice my opinion and therefore will agree	10.53	1.01	0.37
Items displaying DIF in favor of older adults compared to younger adults (older > younger)					
Schizoid	26	Having no interest in sexual encounters with someone else is characteristic of me	10.85	1.03	0.29
Dependent	33	I find it very difficult to disagree in public; even if I totally disagree with	17.90	1.44	0.50

	someone I don't dare to voice my opinion and therefore will agree

Note. L-A LOR=Liu-Agresti common log odds ratio; Cox's B=Cox's Noncentrality parameter estimator.

Table 5

Descriptive Statistics, ANOVA, and Effect Sizes for the ADP-IV PD Scales for the Younger (n=107), Middle-Aged (n=107), and Older (n=107) Adults

Personality Disorder	Cronbach Alpha			Raw Means (*SD*)			Analysis of Variance		Effect Size Cohen's *d*		
	Younger	Middle aged	Older	Younger	Middle aged	Older	*F* (df=2, 320)	*p*-value	1 vs. 2	2 vs. 3	1 vs. 3
Paranoid	.80	.83	.72	19.81 (7.81)	17.18 (7.89)	15.99 (6.52)	7.416	.001*	.35**	.17	.53**
Schizoid	.76	.67	.59	17.56 (7.43)	18.13 (7.16)	20.33 (7.04)	4.391	.013*	-.08	-.31	-.39**
Schizotypal	.80	.79	.66	23.03 (9.04)	21.39 (8.68)	19.93 (7.18)	3.709	.026*	.19	.18	.38**
Antisocial	.66	.58	.50	20.01 (7.56)	14.52 (5.29)	12.69 (4.37)	44.640	.000*	.85**	.38	1.19**
Borderline	.82	.82	.77	33.95 (11.70)	29.64 (10.90)	26.81 (10.17)	11.564	.000*	.38**	.27	.65**
Histrionic	.81	.75	.66	22.01 (8.55)	20.37 (7.96)	17.79 (6.11)	8.378	.000*	.20	.37	.57**
Narcissistic	.77	.78	.75	21.53 (8.01)	19.94 (7.49)	18.46 (6.93)	4.511	.012**	.13	.21	.41**
Avoidant	.84	.86	.78	21.03 (8.82)	19.08 (8.85)	17.83 (7.90)	3.809	.023**	.22	.15	.22

Dependent	.81	.83	.76	21.99 (8.64)	20.42 (9.05)	19.29 (8.04)	2.669	.071	.18	.13	.33
Obsessive-compulsive	.68	.71	.68	23.47 (7.71)	24.52 (8.10)	22.68 (7.76)	1.479	.229	-.13	.23	.10

Note. Effect sizes were based on comparisons of the younger versus middle aged group 1 vs. 2, middle aged versus older adults 2 vs. 3, and younger versus older adults 1 vs. 3; *significant F at $p < .05$;**significant pairwise comparisons at $p < .016$.